Forensic Analysis of Heterogeneous Social Media Data

Aikaterini Nikolaidou, Michalis Lazaridis, Theodoros Semertzidis, Apostolos Axenopoulos
and Petros Daras
Information Technologies Institute, CERTH, Thessaloniki, Greece

Keywords: Social Media Analytics Forensic Platform, Heterogeneous Social Media Data, Ontology, Labeled Property Graph.

Abstract: It is a challenge to aggregate and analyze data from heterogeneous social media sources not only for businesses and organizations but also for Law Enforcement Agencies. The latter's core objectives are to monitor criminal and terrorist related activities and to identify the "key players" in various networks. In this paper, a framework for homogenizing and exploiting data from multiple sources is presented. Moreover, as part of the framework, an ontology that reflects today's social media perceptions is introduced. Data from multiple sources is transformed into a labeled property graph and stored in a graph database in a homogenized way based on the proposed ontology. The result is a cross-source analysis system where end-users can explore different scenarios and draw conclusions through a library of predefined query placeholders that focus on forensic investigation. The framework is evaluated on the Stormfront dataset, a radical right, web community. Finally, the benefits of applying the proposed framework to discover and visualize the relationships between the Stormfront profiles are presented.

1 INTRODUCTION

Social media sites constitute a rich pool of evolving content along with personal data, preferences, activities and relationships. Due to their affordability and accessibility, social media is a means of communication and action for criminal and terrorist organizations. At earlier times, most crimes left breadcrumbs of evidence in the real world. Nowadays, through the interactive social media platforms, offenders engage in illicit practices such as fraud, cyber stalking, cyber bullying etc. (Gambhir, 2018). Terrorists exploit social media to reach audiences for potential recruits, disseminate messages and organize strategic operations (Bertram, 2016). Furthermore, social media play a key role in political socialization in terms of influencing individual behavior and preparedness to participate in collective actions (Passy, 2000).

Law Enforcement Agencies (LEAs) wish to take advantage of these information sources for the sake of security. For monitoring and analyzing criminal-related activity in social media networks, one major question is to identify the most influential profiles, known also as "key player" discovery (Zenou, 2016). LEAs are also interested in answering the so called six W's: Who, What, When, Where, Why and How. These questions are fundamental and are traditionally raised during criminal investigations (Carrier et al., 2003).

Social media evidence provides information about a suspect's or a victim's profile that can be mined in close-to-real-time. The contacts, messages, geolocation data, photos and generally their daily activities are offered in a chronological order. Monitoring and analyzing the abundance of information shared by social media users and social media metadata should theoretically facilitate LEAs to gain insights into mass communication and come to fruitful conclusions in an inquiry. However, this kind of exploration is by no means a straightforward task.

A recent research by Arshad et al. (Arshad et al., 2019) explains the challenges that law enforcement personnel face when handling social media data for forensic investigation. The first issue the authors observe is that in a single social media investigation, some data elements are considered out of context and not taken into account. Moreover, the components of data which they consider important are stored separately. All fragmented and unstructured social media information, although it may seem to be of little importance, would be very useful if it was in a coherent representation and in chronological order. Be-

Nikolaidou, A., Lazaridis, M., Semertzidis, T., Axenopoulos, A. and Daras, P.
Forensic Analysis of Heterogeneous Social Media Data.
DOI: 10.5220/0008347803430350
In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, pages 343-350
ISBN: 978-989-758-382-7

sides, due to the separate storage of data, data analysis is limited to keyword search which is inadequate for a forensic investigation. The second challenge in a social media forensic tool development, is to deal with the heterogeneity of different social media platforms. There are also several references in the literature that emphasize the need for cyber forensic tools that support heterogeneous research in an automated way (Caviglione et al., 2017), (Soltani and Seno, 2017), (Montasari and Hill, 2019). A forensic tool developed for one social media platform, may not be suitable for another platform because they differ in concepts, structure, data formats, and access methods.

According to one of the most accepted approaches for Social Media Analytics (SMA) proposed in (Stieglitz et al., 2014), a typical procedure for SMA consists of the following four steps: i) Discovery of the topics that should be tracked; ii) Tracking, which involves data source selection, approach, method and output; iii) Preparation of the data and iv) Analysis. Regarding analysis step, Karabiyik et al. (Karabiyik et al., 2016) point out that Social Network Analysis (SNA) is crucial during a digital forensic research for understanding the links between profiles. SNA uses graph algorithms to analyze the structure of the network, discovers strategic positions and specific subnetworks. Through SNA, LEAs would discover invaluable hidden associations and knowledge in the social data. Decomposing a social graph to communities yields a deeper insight to these seemingly chaotic interactions. The results from SNA in digital forensic investigations can solve criminal cases, prevent terrorist attacks, track frauds, classify and match social network accounts etc. However, despite the great need to identify and analyze network structures of terrorist and criminal groups, existing digital forensic tools do not provide such functionality.

The objective of this paper is to bridge the gap between the dispersed social media data and the social media forensic applications. This study defines a unified framework to exceed the heterogeneity of different social media platforms and to homogenize their data. The security-oriented framework will permit both the preservation as well as the search and the correlation while providing automated tools for data analysis and visualization. The main goals and contribution of this paper are the following:

- A new ontology is introduced to reflect the majority of today's social media perceptions. Based on this unified ontology, heterogeneous social media data is integrated into a single, structured, cohesive representation. The ultimate goal is to aggregate data not only from social media sites but also from other relevant physical or cyber data sources.
- The unified framework that uses the aforementioned ontology is presented. The framework allows the homogenization of data from multiple sources and thus the consolidated handling and querying of the data for further forensic analysis. It focuses on the three last steps of a SMA process: tracking, preparation and analysis for data from various social media sources.
- The cross-source analysis system, as part of the framework, that supports advanced custom queries and SNA methods for discovery of hidden knowledge. LEAs would use it for investigation and evidence-building through a predefined set of query placeholders that hide all technical complexities.

As a demonstration, the framework is applied to discover and visualize the relationships between the profiles associated with Stormfront[1]. Stormfront is one of many examples of radical right wing virtual communities, but it is the oldest "hate" website.

The implementation is based entirely on open-source tools. The framework's procedures are developed in Python. Choosing an appropriate software architecture and storage technology is one of the major challenges in social media analytics (Stieglitz et al., 2018). Social networks can be very big but sparse since a user is likely to interact with a small percentage of users (Leung and Jiang, 2017). The proposed framework's use case demands a reliable graph storage and retrieval technology that supports efficient execution of complex queries in large and sparse graphs that can scale up to billions of nodes. The labeled property graph database Neo4j [2] fulfills these requirements, and was therefore chosen for this work. The predefined query placeholders have been implemented in Cypher [3] query language. It must be stressed that the proposed ontology at this stage provides the conceptualization of the given domain.

This paper is organized as follows. Related work is reviewed in Section 2. In Section 3 the proposed framework for tracking, preparing and analyzing the social media data is illustrated. Section 4 describes the proposed ontology. Section 5 provides the experimental steps using real data from Stormfront. Section 6 presents the conclusion.

[1] https://www.stormfront.org/forum/index.php

[2] https://neo4j.com/

[3] https://neo4j.com/developer/cypher-query-language/

2 RELATED WORK

Research in social media forensic analysis can be divided into two parts: a) general investigative approaches and b) ontology-based approaches.

There is a significant number of recent studies in the literature that analyze social media data for investigative purposes. In (Mouhssine and Khalid, 2018), a framework is proposed that performs sentiment analysis and detects extremist content on Facebook. A security alert mechanism making use of a supervised machine learning approach for real-time phishing tweet detection on Twitter is presented in (Liew et al., 2019). Alghofaili and Almishari (Alghofaili and Almishari, 2018) detect Arabic-based Twitter profiles that incite terrorism by categorizing accounts as either inciting or non-inciting of terrorism. To do so, they extract relevant features, which are later fed to machine learning-based classifiers to differentiate inciting from non-inciting accounts. Nouh et al. (Nouh et al., 2019) use textual and psycholinguistic signals derived by ISIS for recruitment purposes to detect extremist content on Twitter. An approach for detecting radical users of social network among unknown ones is presented by Petrovskiy and Chikunov (Petrovskiy and Chikunov, 2019). The authors analyse the relationships and features of the users as of vertices of social graph. However according to their method, a part of the network community must be labeled as dangerous, safe or unknown. In (Sánchez-Rebollo et al., 2019), a methodology is developed to detect the leaders orchestrating terrorist networks and their followers that remain partially hidden. They use fuzzy clustering to point out the dangerous profiles and they focus on the analysis of Twitter messages.

The use of ontologies is an approach for modeling social networks that offers several advantages (Oellinger and Oezden Wennerberg, 2006). Masmoudi et al. (Masmoudi et al., 2018) propose a method for mining radicalization indicators from online messages. The social messages are annotated with concepts from a domain ontology the authors designed. The ontology focuses on the users and their messages. During an inference phase the messages exhibiting a radicalization indicator are identified by exploiting the annotations. An ontology named ”SC-Ont” that focuses on data that can be obtained from Facebook and Google is published by Kalemi and Yildirim-Yayilgan (Kalemi and Yildirim-Yayilgan, 2016). The ontology serves for cataloging online information regarding suspects and information from witnesses. Tserpes et al. (Tserpes et al., 2012) developed an ontology called ”SocIoS” on top of the concept that entities in the various online social network sites have largely similar notions. Their ontology is an effort to cover most of the concepts of the social networks. However, the model does not cover some today's important concepts and it is not explained how to use it. Fang et al. (Fang et al., 2019) also proposed an ontology to aggregate data from different social network sites for event analysis purposes. The design of a specific analysis application is out of the scope of their work.

In this paper we expand on the previous work in (Tserpes et al., 2012) and propose an ontology to cover the concepts of modern social media platforms. Social media data and interactive behaviours common to heterogeneous social media platforms will be stored on the basis of the ontology in a structured and coherent way in a common graph database. The result will be a sophisticated security-oriented cross-source forensic analysis application. LEAs will have the opportunity to explore all aspects of the information provided by the social media sources. The investigators do not need to have technical knowledge since a set of predefined query placeholders related to forensic inquiry will be at their disposal.

3 PROPOSED FRAMEWORK

The proposed framework is depicted in Figure 1, and mainly corresponds to the three last steps of a Social Media Analytics process: tracking, preparation and analysis for data from multiple social media sources.

After data tracking by platform-specific APIs provided by online social networks such as Facebook, Twitter, LinkedIn and others, or crawlers, the preparation phase follows. In this work, a new ontology is introduced to reflect the majority of today's social media perceptions. The new ontology will be called Unified Social Network Ontology (USNO). To easily adapt the proposed approach, social-media-specific plugins have been developed for some social media sites: Skype, Twitter, Stormfront. A separate plugin can be developed for each social media site of interest. Each social-media-specific plugin, takes as input the corresponding social media data. The USNO describes in detail how the social media data must be stored in the graph database. It depicts the node labels, the node properties and the relationships that will link the nodes/entities in the graph. The data is processed according to each social media site to retrieve implicit knowledge related to the unified ontology we introduced. Then, the data is mapped to the USNO.

In the next step, the social media data is stored in the graph database Neo4j. For each element a new node is created. For example a ”hashtag” shared by multiple posts is represented by a single node. In order to thoroughly analyze the interaction between the profiles, an inference mechanism has been implemented to create direct links between the profiles that communicate with each other if these links are not explicitly provided. Also a mechanism to detect data inconsistencies in the graph database is supported. For example, we developed a Cypher rule to check the validity of the profile identification numbers. A profile identification number must correspond to a unique username for a single social media platform.

As for the analysis phase, we have developed Cypher queries to retrieve information and relationships related to forensic research on profiles. Queries can take as arguments the inputs of the end-users. For example the objective of a query could be the following: Given the location of a crime and a time frame, return all conversations and conversationalists, events and media items that refer to the crime location during this time frame. Furthermore, SNA can be performed thanks to an efficiently implemented library of common graph algorithms exposed as Cypher procedures provided by Neo4j (Amy E. Hodler, 2019). They can provide insights on relevant profiles using centrality algorithms, or implicit structures like communities using community detection algorithms. The result after the algorithms execution can be written back as a node property and be used for further analysis. Investigators have the opportunity to perform complex data analysis and visualize the results, at the touch of a button. More examples on the analysis phase are provided in section 5.

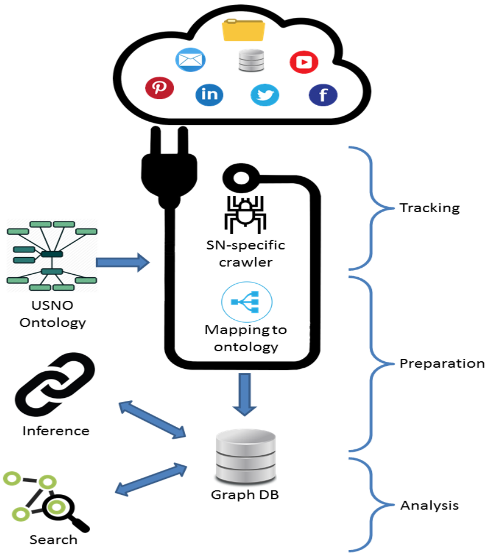

Figure 1: Proposed framework for data preparation and analysis.

4 UNIFIED SOCIAL NETWORK ONTOLOGY

Unified Social Network Ontology is depicted in Figure 2. It consists of 11 classes that are associated with 20 relationships. In the following section, the classes along with their properties are presented.

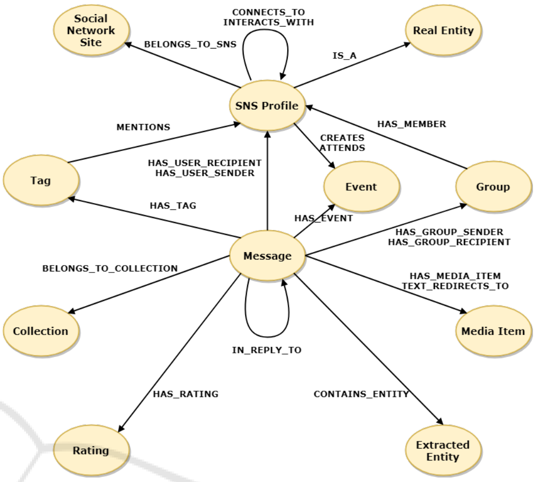

Figure 2: Unified Social Network Ontology.

4.1 Classes and Properties

In addition to the attributes listed below, each class has another property, its identification number.

1. Social_Network_Site: It can be any social media site like Facebook, Instagram, Stormfront etc. The properties include:
   - Platform: The name of the social network platform e.g. Facebook, Instagram, Stormfront etc.
2. SNS_Profile: A social network profile. The properties include:
   - Username: A profile’s username.
   - Profile_title: A profile’s title e.g. member, guest.
   - Email: A profile’s email address.
   - Profile_lang: The speaking language of a profile.
   - Registered_date: The date a profile became member of the social network.
   - Ethnicity: The ethnicity of the profile.
   - Gender: The gender of the profile.
   - Birthday: The birthday of the profile.
3. Real_Entity: A real physical entity e.g. a person, a business, a group of people. The properties include:

- Entity_type: The real physical person(s) who manage the profile e.g. person, business, group of people.

4. Message: Refers either to inbox or post. The properties include:
   - Msg_type: The type of the message e.g. post or inbox.
   - Content: The full content of the message.
   - Msg_lang: The language in which the message was written.
   - Msg_security: The security level of the message e.g. public or private.
   - Date_sent: The date and time of sharing the message.
   - Num_views: The number of views of the message.
5. MediaItem: A media item can have different content formats such as embedded image, embedded video, youtube video, hyperlink with clicking text or image and it is attached to a message.
   - Mediaitem_type: The type of the media item e.g. embedded image, embedded video, youtube video, hyperlink with clicking text or image.
   - Mediaitem_url: The url of the media item.
   - Mediaitem_text: The alternate text of an image or the content of a clicking text.
   - Filename: The filename of the media item.
   - Lng_lat: A text representation of the coordinate data as latitude and longitude e.g. ”470999 1234300”.
6. Collection: A collection contains messages with common elements e.g. topic, hashtag.
   - Collection_type: The collection’s type e.g. topic, library, shared folder, hashtag.
   - Collection_name: The collection’s name.
7. Group: A group provides a space to communicate about shared interests with certain people.
   - Group_title: Title of the group.
   - Description: Description of the group.
   - Group_security: The security level of the group e.g. public, closed or secret.
8. Tag: A link to a profile along with a media item.
   - Bounding_box: A text representation of 2 longitudes and 2 latitudes coordinates of a tagged profile in a media item e.g ”6.73462 6.75652 -53.57835 -53.56289”.
9. Event: An event lets a profile to organize and respond to gatherings in the real world.
   - Event_title: The event’s title.
   - Event_location: An abstract region of space (e.g. a geospatial point or region) where the event will take place.
   - Event_security: The security level of the event e.g. private or public.
   - Event_timestamp: The time and date of the event.
10. Rating: A measurement of how good or popular a message is.
    - Num_like: The score that depicts the positive rating of a message.
    - Num_dislike: The score that depicts the negative rating of a message.
11. Extracted_Entity: An entity that is extracted from a message after performing named-entity recognition. It can be: location, person or organization.
    - Extracted_entity_type: The extracted entity’s type e.g. location, person or organization
    - Extracted_entity_name: The extracted entity’s name. e.g. for type ”location” the name could be ”London”.

5 STORMFRONT CASE STUDY

To demonstrate our framework we have implemented a plugin for Stormfront, the world’s oldest and largest white supremacist internet forum, which has hosted some of the most lethal hate criminals and mass murderers of the right-wing extremist movement. Analysis for Stormfront data has gained a lot of attention. According to a two-year investigation by Southern Poverty Law Center in 2014, registered Stormfront users have been responsible for approximately 100 murders (Beirich, 2014). The proposed model is instantiated on Stormfront social case study to be validated. Extensive Stormfront data collected by a crawler are used, for the period from September 2001 to July 2017 totaling 4,488,476 posts and 355008 users. We will present how LEAs could explore this dataset.

5.1 Stormfront Data Preparation

Data was initially in a raw form, thus it was very difficult to analyze and discover the ”key players”. In the preparation phase, the ”Stormfront plugin”, takes as input the data and analyzes the contents of the posts to discover implicit information related to USNO. For example the reply posts, the posts with media files, the images and texts with hyperlinks are retrieved.

Also named-entity recognition is supported. Every post message is examined for named entity mentions of type person, location and organization. After this stage, the data is mapped to the USNO and stored in the Neo4j. The ”Stormfront plugin”, creates all the nodes, properties and relationships in the graph database.

To directly associate profiles and perform SNA to retrieve hidden knowledge, a Cypher rule has been developed to create a new ”INTERACTS_WITH” relationship between profiles that have communication interaction. For example, profile ”A” ”INTERACTS_WITH” profile ”B” if ”A” has responded to a post shared by ”B”, thus directly linking the two profiles. In order to detect potential inappropriate profile identification numbers, a Cypher rule checks that each profile identification number corresponds to a unique username. Figure 3 illustrates the Stromfront Neo4j schema.

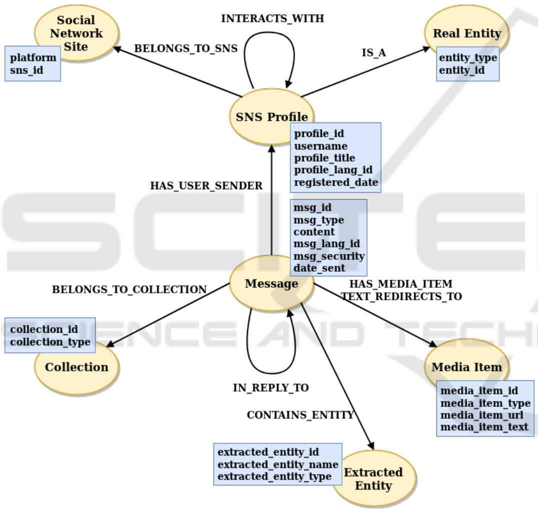

Figure 3: Stormfront Neo4j schema.

5.2 Stormfront Data Analysis

Neo4j provides a library of common graph algorithms exposed as Cypher procedures. Graph algorithms in this case could be applied to the subgraph containing nodes labeled ”SNS_Profile” and associated with the relationship ”INTERACTS_WITH”. The interaction between social network profiles provides valuable information regarding their importance and respective role within the subgraph. Since most graph algorithms run better on a connected component, initially the investigators could isolate the largest subgraph(s) of profiles that interact with each other to understand a graph’s structure. The Connected Components algorithm is applied behind the scenes. After that, the ”islands” will be revealed where each node is accessible from any other node in the same set.

To discover communities in the ”islands” community detection algorithms would be applied. Tsvetovat and Kouznetsov (Tsvetovat and Kouznetsov, 2011) report that Al Qaeda members are organized in small cells during training and preparation for terrorist attacks. The cells have a dense triadic structure, with everyone embedded in triads with everyone else. LEAs could use the cross-source analysis system to discover the triangles in the network. A triangle is a set of three nodes, where each node has a relationship to all other nodes. They could also discover the number and density of the triangles to which a profile is a member. In these cases, the Triangle Counting and Clustering Coefficient algorithms are applied. Investigators could also detect communities in the ”islands” using the Louvain Modularity algorithm. Communities are groups of nodes within a network that are more densely connected to one another than to other nodes.

In order to detect the ”key players”, their role and their impact on the network, centrality algorithms would be used. To quantify the amount of influence a node has over the flow of information in the graph, the Betweenness Centrality algorithm is applied. Betweenness is appropriate to discover the most central profiles, nodes that serve as a bridge from one part of a graph to another. Closeness Centrality algorithm is applied to detect the profiles that are able to spread information very efficiently through the graph. To measure the transitive influence of the profiles the PageRank algorithm is used. This is a very important centrality measure as it takes into account the influence of a node’s neighbors, and their neighbors.

Querying the graph with Cypher gives important insights to the discovery of important profiles, and information about them and their connection to other profiles. The queries can focus on a particular known profile. Below are listed some examples of automated queries provided by our platform:

- Given specific keywords, discover the profile that sent the most messages that started a conversation related to the keywords.
- Given a specific profile, list the profiles that he/she will engage and respond.
- Discover the length of the conversations and the number of people who are involved in them.
- Given a specific profile, discover the number of conversations started by him/her, the average length of the conversations and the average participants.
- Track conversations about a particular topic or for a specific period of time.

- Discover post volume by day or month.
- Explore around certain entities of type person, location and organization. For example, discover which locations are mentioned for a specific date.

Our tool gives investigators the opportunity to answer very complex custom queries at a touch of a button. For example, the following query runs behind the scenes to detect ”cliques”, that is, groups of profiles that post to common topics. The investigators would only give the input parameters. In this case the parameters are ”number_of_common_topics”, ”from_period” and ”to_period”.

```
MATCH
(u1:SNS_Profile)<-[:HAS_USER_SENDER]
-(m:Message)
-[:BELONGS_TO_COLLECTION]->(c:Collection)
<-[:BELONGS_TO_COLLECTION]-(m1:Message)
-[:HAS_USER_SENDER]->(u:SNS_Profile)
WHERE ID(u1) > ID(u)
AND m.date_sent > $from_period
AND m.date_sent < $to_period
AND m1.date_sent > $from_period
AND m1.date_sent < $to_period
WITH collect(DISTINCT c) AS topics,
u1.username AS user, u.username AS users
WHERE
size(topics)>$number_of_common_topics
WITH collect(users) AS group,topics,user
WITH collect(user)+group AS clique,
size(topics) AS number_of_topics
RETURN clique, number_of_topics
ORDER BY number_of_topics DESC
```

A similar but more concrete query could return the cliques, where a specific ”targeted” profile takes part, revealing profiles in his/her circle.

6 CONCLUSIONS

In this paper, we aim at homogenizing data from multiple social media sources to bring them into a unified cross-source analysis system for forensic purposes. The Unified Social Network Ontology has been introduced as part of a framework for tracking, preparing and analyzing data. The ontology covers the majority of today's social media perceptions. We have developed social-media-specific plugins to map the data to the USNO and store them in a labeled property graph database. As a result the raw data are associated and stored in a coherent representation based on the USNO. The homogeneity of the data and its retention in a common database is the basis on which we have implemented a sophisticated forensic tool to go beyond keyword search and respond to complicated investigation queries. We designed a number of predefined query placeholders which give insights into the network structure and reveal the most important profiles through SNA methods. In addition, LEAs will easily detect any indication that could provide the social media data. For example law enforcement personnel will track conversations and their frequency, discover all information about a specific profile, explore around a particular topic for a given location and time frame, and so on. The automatic data homogenization, preservation and analysis will reduce significantly the investigation cost.

In the demonstration and case study, the proposed framework has been applied to Stormfront internet forum. The plugin we developed for Stormfront is responsible for mapping the data to the USNO and importing them into the graph database. After this stage, we demonstrated how LEAs could explore the graph to discover important profiles, relationships and evidence through the library of predefined query placeholders. An advantage of the proposed framework is that it can be easily adapted to represent data from other social networks. As a proof of concept, it has also been applied to Twitter dataset[4] to investigate Twitter-related data like mentions, hashtags and retweets.

In future work, we will investigate new analysis algorithms such as profile matching to be added to the library with the predefined query placeholders. Thus, a new relationship ”MATCHES_WITH” will be created between the profiles that correspond to the same person. Associating user profiles to many social networks would provide, among other things, the opportunity to better understand the interplay between the different types of suspect's operations.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by the European Commission under contract H2020-700381 ASGARD.

REFERENCES

Alghofaili, H. and Almishari, M. (2018). Countering terrorism incitement of twitter profiles in arabic-context. In *2018 21st Saudi Computer Society National Computer Conference (NCC)*, pages 1–5.

Amy E. Hodler, M. N. (2019). *Graph Algorithms: Practical Examples in Apache Spark and Neo4j.* ” O'Reilly Media, Inc.”.

[4] https://www.kaggle.com/fifthtribe/how-isis-uses-twitter

Arshad, H., Jantan, A., and Omolara, E. (2019). Evidence collection and forensics on social networks: Research challenges and directions. *Digital Investigation*.

Beirich, H. (2014). White homicide worldwide. *SPLC Intelligence Report*, (154).

Bertram, L. (2016). Terrorism, the internet and the social media advantage: Exploring how terrorist organizations exploit aspects of the internet, social media and how these same platforms could be used to counter-violent extremism. *Journal for Deradicalization*, (7):225–252.

Carrier, B., Spafford, E. H., et al. (2003). Getting physical with the digital investigation process. *International Journal of digital evidence*, 2(2):1–20.

Caviglione, L., Wendzel, S., and Mazurczyk, W. (2017). The future of digital forensics: Challenges and the road ahead. *IEEE Security Privacy*, 15(6):12–17.

Fang, M., Li, Y., Hu, Y., Mao, S., and Shi, P. (2019). A unified semantic model for cross-media events analysis in online social networks. *IEEE Access*.

Gambhir, A. (2018). Cyber crime and their impacts in various aspects. *International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)*, 4.

Kalemi, E. and Yildirim-Yayilgan, S. (2016). Ontologies for social media digital evidence. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 10(2):335–340.

Karabiyik, U., Canbaz, M. A., Aksoy, A., Tuna, T., Akbas, E., Gonen, B., and Aygun, R. (2016). A survey of social network forensics. *Journal of Digital Forensics, Security and Law (JDFSL)*, 11:55–128.

Leung, C. K. and Jiang, F. (2017). Efficient mining of'following'patterns from very big but sparse social networks. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 1025–1032. ACM.

Liew, S. W., Sani, N. F. M., Abdullah, M. T., Yaakob, R., and Sharum, M. Y. (2019). An effective security alert mechanism for real-time phishing tweet detection on twitter. *Computers & Security*, 83:201–207.

Masmoudi, A., Barhamgi, M., Faci, N., Saoud, Z., Belhajjame, K., Benslimane, D., and Camacho, D. (2018). An ontology-based approach for mining radicalization indicators from online messages. In *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, pages 609–616.

Montasari, R. and Hill, R. (2019). Next-generation digital forensics: Challenges and future paradigms. In *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, pages 205–212.

Mouhssine, E. and Khalid, C. (2018). Social big data mining framework for extremist content detection in social networks. In *2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, pages 1–5.

Nouh, M., Nurse, J. R. C., and Goldsmith, M. (2019). Understanding the radical mind: Identifying signals to detect extremist content on twitter.

Oellinger, T. and Oezden Wennerberg, P. (2006). Ontology based modeling and visualization of social networks for the web. volume 2, pages 489–497.

Passy, F. (2000). Socialization, recruitment, and the structure/agency gap. a specification of the impact of networks on participation in social movements. In *Social Movement Analysis: The Network Perspective, a workshop held at Ross Priory, Loch Lomond, Scotland, June*, pages 22–25.

Petrovskiy, M. and Chikunov, M. (2019). Online extremism discovering through social network structure analysis. In *2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT)*, pages 243–249.

Sánchez-Rebollo, C., Puente, C., Palacios, R., Piriz, C., Fuentes, J. P., and Jarauta, J. (2019). Detection of jihadism in social networks using big data techniques supported by graphs and fuzzy clustering. *Complexity*, 2019.

Soltani, S. and Seno, S. A. H. (2017). A survey on digital evidence collection and analysis. In *2017 7th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 247–253.

Stieglitz, S., Dang-Xuan, L., Bruns, A., and Neuberger, C. (2014). Social media analytics - an interdisciplinary approach and its implications for information systems. *Business & Information Systems Engineering*, 6:89–96.

Stieglitz, S., Mirbabaie, M., Ross, B., and Neuberger, C. (2018). Social media analytics–challenges in topic discovery, data collection, and data preparation. *International journal of information management*, 39:156–168.

Tserpes, K., Papadakis, G., Kardara, M., Papaoikonomou, A., Aisopos, F., Sardis, E., and Varvarigou, T. A. (2012). An ontology for social networking sites interoperability. In *KEOD*, pages 245–250.

Tsvetovat, M. and Kouznetsov, A. (2011). *Social Network Analysis for Startups: Finding connections on the social web.* " O'Reilly Media, Inc.".

Zenou, Y. (2016). Key players. *Oxford Handbook on the Economics of Networks, Oxford: Oxford University Press, forthcoming*.